Internship Project Report

Internship Role: Data Analysis Intern **Company:** Nullclass **Project Title:** Google Play Store Data Analysis

1. Introduction

During my internship at Nullclass, I worked on analyzing **Google Play Store app performance** using advanced data visualization techniques. The project aimed to develop an interactive dashboard that helps stakeholders understand key trends related to **installs, revenue, app ratings, and reviews**, allowing informed decision-making.

2. Background

With millions of apps available on the Play Store, **app developers and marketers** need actionable insights to optimize their products. This internship project involved exploring **data relationships, correlations, and trends** among various app categories to aid business strategy and growth.

3. Learning Objectives

- Understand **data cleaning, filtering, and visualization** techniques.
- Develop **interactive dashboards** with conditional display features.
- Enhance **Python, Pandas, and Plotly skills** for data analysis.
- Explore **user engagement trends** and monetization insights.

4. Activities and Tasks

Throughout the internship, I worked on multiple **visualizations**, each designed to highlight specific aspects of Play Store data:

Task 1: Scatter Plot (Revenue vs. Installs)

- Created a **scatter plot** showing the relationship between **app installs and revenue** for paid apps.
- Implemented a **trendline** to analyze correlation and **color-coded data points** by app category.

Task 2: Choropleth Map (Global Installs)

- Built an interactive **global install visualization** using Plotly.
- Filtered data to show only the **top 5 categories** and **excluded specific app categories** based on name conditions.
- Applied **visibility constraints** so the map appeared **only between 6 PM and 8 PM IST**.

Task 3: Grouped Bar Chart (Ratings & Reviews)

- Compared **average rating and total review count** for **top 10 app categories** by installs.
- Excluded categories with **ratings below 4.0**, app sizes **below 10 MB**, and **last updated before January**.
- Applied **dashboard visibility restriction** between **3 PM and 5 PM IST**.

Task 4: Dual-Axis Chart (Free vs. Paid Apps)

- Compared **installs and revenue trends** for free vs. paid apps in the **top 3 categories**.
- Applied various **filters** for app size, content rating, Android version, and app name restrictions.
- Ensured visibility was **limited between 1 PM and 2 PM IST**.

Task 5: Bubble Chart (App Size vs. Rating)

- Analyzed how **app size correlates with ratings**, with **bubble size representing installs**.
- Applied filters to include only **apps with a rating >3.5** in the **Games** category and **installs >50K**.
- Made sure the graph **only appeared between 5 PM and 7 PM IST**.

Task 6: Heatmap (Correlation Analysis)

- Generated a **heatmap matrix** showing relationships between installs, ratings, and review counts.
- Filtered data based on **recent updates, minimum installs, and genre name restrictions**.

- The visualization was designed to be **visible only between 2 PM and 4 PM IST**.

Task 7: Violin Plot (Rating Distribution by Category)

- Developed a **violin plot** to visualize rating distributions **only for categories with >50 apps**.
- Applied constraints ensuring apps contained the **letter "C" in their name**, had **>10 reviews**, and **ratings below 4.0**.
- Ensured the plot was **visible between 4 PM and 6 PM IST**.

5. Skills and Competencies Acquired

- **Data Analysis:** Cleaning and filtering large datasets efficiently.
- **Visualization Expertise:** Creating **interactive and insightful charts** using Python libraries.
- **Technical Skills:** Python (Pandas, Plotly), dashboard development, and **conditional visibility implementation**.
- **Problem-Solving:** Tackling challenges related to large-scale data processing.

6. Feedback and Evidence

- The **interactive dashboard improved data accessibility** for decision-makers at Nullclass.
- Conditional filters **ensured relevant visualizations appeared only within designated time frames**, optimizing analysis efficiency.
- My mentor and team provided **positive feedback** on the project's clarity, usability, and insights.

7. Challenges and Solutions

- **Challenge:** Restricting graphs based on time constraints.
    - **Solution:** Used **Dash framework** for **real-time visibility conditions** in the dashboard.
- **Challenge:** Filtering large datasets efficiently.
    - **Solution:** Applied optimized **query processing** and indexing techniques.
- **Challenge:** Ensuring meaningful category-based insights.
    - **Solution:** Iteratively **refined filtering logic** based on team feedback.

8. Outcomes and Impact

The project successfully enhanced the **data analytics dashboard** by incorporating meaningful insights and improving overall **data-driven decision-making** for app developers and marketers.

9. Conclusion

This internship was an invaluable experience that strengthened my **technical, analytical, and problem-solving skills**. It provided exposure to **real-world business challenges** and equipped me with practical expertise in **data-driven decision-making**.